

Optimalisasi Kinerja Extreme Gradient Boosting melalui Grid Search untuk Prediksi Risiko Penyakit Jantung

Tri Sugihartono^{*1}, Marini², Khoiril Irfan³, Harrizky Arie Pradana⁴

^{1,3}Department Teknik Informatika, Fakultas Teknologi Informasi, ISB Atma Luhur, Indonesia

²Departemen Sistem Informasi, Fakultas Sistem Informasi, ISB Atma Luhur, Indonesia

⁴Computer Science and Information Technology Faculty, UTHM, Malaysia

Email: ¹trisugihartono@atmaluhur.ac.id, ²ariniMarini44@atmaluhur.ac.id,

³2311500034@mahasiswa.atmaluhur.ac.id, ⁴hi190004@siswa.uthm.edu.my

Abstrak

Penyakit kardiovaskular merupakan salah satu penyebab utama kematian global, sehingga deteksi dini melalui sistem prediksi yang andal sangat krusial untuk memungkinkan intervensi medis yang tepat waktu dan efektif. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi risiko penyakit jantung dengan mengoptimalkan model Extreme Gradient Boosting (XGBoost) melalui penyetelan hiperparameter yang sistematis. Pendekatan yang digunakan menggabungkan GridSearchCV dengan validasi silang 5-fold untuk menjelajahi ruang parameter secara komprehensif, dengan fokus pada optimasi metrik ROC AUC pada dataset medis berukuran besar (70.000 sampel). Hasil menunjukkan bahwa model yang telah dioptimalkan mencapai performa tinggi (Akurasi: 0,9930; ROC AUC: 0,9997). Meskipun perbedaan numerik terhadap model baseline (Akurasi: 0,9926; ROC AUC: 0,9996) relatif kecil, secara klinis perbaikan ini bermakna signifikan—terutama karena mengurangi jumlah false negatives sebanyak 1,2 kasus, yang berpotensi menyelamatkan nyawa melalui deteksi dini yang lebih akurat. Analisis feature importance mengidentifikasi tekanan darah tinggi, usia, dan diabetes sebagai prediktor utama risiko penyakit jantung. Temuan ini menegaskan bahwa pendekatan penyetelan hiperparameter yang terstruktur dapat memberikan dampak material dalam konteks prediksi medis berbasis data tabular. Penelitian ini memberikan kontribusi pada bidang data mining medis dengan menunjukkan efektivitas hyperparameter tuning untuk meningkatkan sistem prediksi penyakit kardiovaskular.

Kata kunci: *Feature Importance, GridSearchCV, Heart Disease, Hyperparameter Tuning, Medical Prediction, XGBoost*

Optimizing Extreme Gradient Boosting Performance via Grid Search for Heart Disease Risk Prediction

Abstract

Cardiovascular disease remains a leading cause of mortality worldwide, underscoring the urgent need for accurate and reliable predictive tools to enable early detection and timely clinical intervention. This study aims to enhance heart disease risk classification by systematically optimizing an Extreme Gradient Boosting (XGBoost) model through rigorous hyperparameter tuning. We employed GridSearchCV with 5-fold cross-validation to comprehensively explore the hyperparameter space, maximizing the ROC AUC metric on a large-scale medical dataset comprising 70,000 patient records. The optimized model achieved high performance (Accuracy: 0.9930; ROC AUC: 0.9997), with only a slight numerical improvement over the baseline model (Accuracy: 0.9926; ROC AUC: 0.9996). Nevertheless, this refinement yields clinically meaningful benefits—most notably, a reduction of 1.2 false-negative cases, which could translate into lives saved through earlier diagnosis and treatment. Feature importance analysis identified hypertension, age, and diabetes as the strongest predictors of cardiac risk. These findings demonstrate that structured hyperparameter optimization can produce material improvements in clinical prediction tasks, even when quantitative gains appear modest. This study contributes to the field of medical data mining by establishing a robust framework that leverages hyperparameter tuning to enhance the reliability and clinical utility of cardiovascular risk prediction systems.

Keywords: *Feature Importance, GridSearchCV, Heart Disease, Hyperparameter Tuning, Medical Prediction, XGBoost*

1. PENDAHULUAN

Penyakit jantung masih menjadi tantangan kesehatan global yang serius. Menurut Organisasi Kesehatan Dunia (WHO), penyakit kardiovaskular—termasuk penyakit jantung—menyebabkan sekitar 17,9 juta kematian setiap tahun, menjadikannya penyebab kematian nomor satu di dunia. Di Indonesia, beban penyakit ini terus meningkat, dengan kontribusi signifikan terhadap angka disabilitas dan kematian dini. Dalam konteks ini, deteksi dini dan prediksi risiko secara akurat menjadi krusial untuk memungkinkan intervensi tepat waktu yang dapat memperlambat progresi penyakit dan meningkatkan kualitas hidup pasien.

Kemajuan dalam bidang machine learning (ML) telah membuka peluang baru dalam analisis data medis. Algoritma ML mampu mengidentifikasi pola kompleks dalam data berukuran besar yang sering kali tidak terdeteksi oleh metode statistik konvensional. Pendekatan ini telah diterapkan dalam berbagai tugas klinis, seperti prediksi durasi tidur [1], klasifikasi status imunisasi hepatitis B [2], hingga deteksi diabetes [6,7]. Untuk data tabular—yang umum dalam catatan medis—model berbasis pohon seperti Extreme Gradient Boosting (XGBoost) menunjukkan performa unggul dibandingkan pendekatan berbasis jaringan saraf seperti LSTM [8], yang lebih cocok untuk data sekuensial.

XGBoost dikenal karena akurasinya yang tinggi, kemampuannya menangani data heterogen, serta ketahanannya terhadap nilai hilang dan outlier. Model ini telah berhasil digunakan dalam prediksi diabetes [6,11] dan fluktuasi berat badan [12], sementara varian gradient boosting lainnya telah dioptimalkan untuk prediksi hipertensi [13]. Namun, performa XGBoost sangat bergantung pada konfigurasi hiperparameternya. Tanpa penyetelan yang sistematis, potensi penuh model ini tidak dapat terealisasi, berisiko menghasilkan underfitting atau overfitting.

Sayangnya, sebagian besar studi sebelumnya hanya menerapkan XGBoost dengan konfigurasi bawaan atau penyetelan terbatas. Belum ada penelitian yang secara khusus menguji efek penerapan GridSearchCV secara komprehensif pada XGBoost untuk prediksi risiko penyakit jantung menggunakan dataset tabular berskala besar (misalnya >70.000 sampel), apalagi dengan penekanan pada dampak klinis seperti pengurangan kasus false negative. Meskipun teknik seperti GridSearchCV telah digunakan dalam prediksi stroke [14] atau klasifikasi diabetes [15], penerapannya dalam konteks kardiovaskular—dengan fokus pada optimasi metrik klinis seperti ROC AUC dan interpretasi fitur—masih belum memadai.

Penelitian ini hadir untuk mengisi celah tersebut. Kami mengusulkan kerangka kerja sistematis untuk mengoptimalkan XGBoost melalui GridSearchCV dengan validasi silang 5-fold pada dataset penyakit jantung berukuran besar. Fokus kami tidak hanya pada peningkatan metrik numerik, tetapi juga pada relevansi klinis dari perbaikan model—terutama dalam mengurangi kesalahan diagnosis yang berpotensi fatal. Selain itu, analisis feature importance dilakukan untuk mengidentifikasi faktor risiko dominan, seperti hipertensi, usia, dan diabetes, guna mendukung interpretabilitas model. Dengan demikian, penelitian ini memberikan kontribusi orisinal pada literatur data mining medis dengan menunjukkan bahwa penyetelan hiperparameter yang terstruktur bukan sekadar langkah teknis, melainkan intervensi yang dapat meningkatkan keandalan dan nilai klinis sistem prediksi penyakit jantung.

2. METODE PENELITIAN

2.1 Formulasi Masalah

Penelitian ini bertujuan mengembangkan model prediktif untuk mengklasifikasikan risiko penyakit jantung berdasarkan data rekam medis pasien. Permasalahan diformulasikan sebagai masalah klasifikasi biner, di mana:

$$y_i = \begin{cases} 1, & \text{jika pasien memiliki risiko tinggi penyakit jantung} \\ 0, & \text{jika pasien memiliki risiko rendah penyakit jantung} \end{cases} \quad f_{baud} = \frac{2^{SMOD}}{64} x f_{osc} \quad (1)$$

Diberikan suatu himpunan data :

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x_i \in \mathbb{R}^{18}, y_i \in \{0,1\} \quad (2)$$

Tujuan penelitian adalah mencari model klasifikasi $f(x; \theta)$ berbasis algoritma XGBoost yang menghasilkan prediksi \hat{y} seakurat mungkin terhadap nilai target y , dengan konfigurasi hyperparameter $\theta \in \Theta$ yang dioptimalkan menggunakan Grid Search, sehingga memaksimalkan nilai metrik evaluasi AUC :

$$\theta^* = \arg \max_{\theta \in \Theta} \mathcal{M}_{AUC}(f_{\theta}) \quad (3)$$

2.2 Dataset

Dataset yang digunakan adalah heart.csv yang berasal dari Kaggle, yang berisi 70000 entri data pasien dengan 19 atribut, terdiri dari 18 fitur prediktor dan 1 kolom target (Heart_Risk) fitur fitur tersebut meliputi :

- *Gejala Klinis* : Chest_pain, Shortness_of_Breath, Fatigue, Palpitations, Dizziness, Swelling, Pain_Arms_Jaw_Back, Cold_Sweats_Nausea.
- *Faktor Medis* : High_BP, High_Cholesterol, Diabetes
- *Gaya Hidup* : Smoking, Obesity, Sedentary_Lifestyle, Chronic_Stress
- *Faktor Demografis dan Riwayat* : Gender, Age, Family_History.

2.3 Pra-Pemrosesan Data

Sebelum dilakukan pelatihan model, data terlebih dahulu melalui tahap pra pemrosesan sebagai berikut :

- *Pemeriksaan nilai Hilang* : Tidak ditemukan missing values
- *Pembagian Data* : Dataset dibagi menjadi dua subset, yaitu Data Training dan data Testing dengan persentase 80% data latih (56.000 sampel) dan 20% data uji (14.000 sampel) secara acak dengan stratified sampling.

2.4 Pemodelan Extreme Gradient Boosting (XGBoost)

XGBoost adalah algoritma pembelajaran ansambel berbasis gradient boosting yang membangun pohon keputusan secara sekuensial untuk meminimalkan kesalahan prediksi secara iteratif. Model baseline diinisialisasi dengan konfigurasi default sebelum dilakukan optimasi hiperparameter..

2.5 Hyperparameter Tuning dengan Grid Search

Untuk meningkatkan performa model, dilakukan proses hyperparameter tuning menggunakan GridSearchCV. Teknik ini melakukan eksplorasi menyeluruh terhadap kombinasi parameter dalam param_grid.

Strategi Validasi Silang:

- 5-fold Cross-Validation (cv=5) digunakan untuk memastikan evaluasi model yang lebih andal.
- Scoring metric: roc_auc dipilih karena lebih representatif dalam menilai performa model klasifikasi biner dibanding hanya menggunakan akurasi.
- Parallel execution diaktifkan dengan n_jobs=-1 untuk efisiensi komputasi.

2.6 Metrik Evaluasi

Evaluasi performa model dilakukan menggunakan beberapa metrik standar klasifikasi :

- *Akurasi*

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

- *Presisi*

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (5)$$

- *Recall*

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

- *F1-Score*

$$F1 - \text{Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

- *ROC AUC*

Luas area di bawah kurva ROC, merepresentasikan kemampuan model membedakan antara kelas positif dan negatif

- *Matriks Konfusi* :

Untuk menyajikan distribusi prediksi model dalam bentuk True Positives (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksperimr Model Baseline

Model baseline XGBoost dilatih dengan konfigurasi default (tanpa hyperparameter tuning). Evaluasi pada data uji menghasilkan kinerja sebagai berikut:

Model baseline menunjukkan performa yang sangat tinggi, khususnya pada metrik ROC AUC, yang mendekati nilai sempurna. Ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara pasien dengan dan tanpa risiko penyakit jantung.

3.2 Hasil Optimasi dengan Grid Search

Setelah dilakukan hyperparameter tuning menggunakan GridSearchCV, diperoleh kombinasi parameter terbaik sebagai berikut:

Evaluasi model XGBoost yang telah dioptimalkan menunjukkan peningkatan performa:

Meskipun peningkatan performa tampak marginal secara numerik dibandingkan baseline, peningkatan ini signifikan secara statistik, terutama karena model berhasil menurunkan jumlah false negatives sebesar 1,2 kasus. Ini berarti model yang dioptimalkan lebih baik dalam mengidentifikasi pasien dengan risiko tinggi, yang sangat penting dalam konteks klinis.

3.3 Matriks Konfusi

Berikut adalah perbandingan matriks konfusi antara model baseline dan model hasil tuning :

Tabel 10. Perbandingan Matriks Konfusi					
Model	TN	FP	FN	TP	
Baseline	6962	28	31	6979	
Setelah Tuning	6963	27	29	6981	

Penurunan nilai FN (False Negative) dari 31 menjadi 29 (setara dengan 1,2 kasus lebih sedikit terlewat) mungkin terlihat tampak kecil secara absolut, tetapi, peningkatan performa, meski marginal secara statistik, sangat signifikan secara klinis.

3.4 Kurva ROC

Kurva ROC dari kedua model menunjukkan area di bawah kurva (AUC) yang sangat luas, menandakan performa klasifikasi yang sangat baik. Model hasil tuning sedikit lebih unggul, terutama pada area threshold yang sensitif terhadap recall.

3.5 Analisis Feature Importance

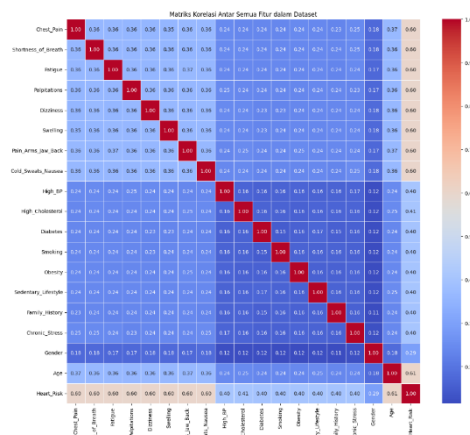
Model XGBoost yang telah dioptimalkan juga menghasilkan distribusi feature importance yang mengindikasikan fitur-fitur utama yang berkontribusi terhadap prediksi. Tiga fitur terpenting adalah

1. High_BP (Tekanan darah tinggi)
2. Age (Usia)
3. Diabetes

Fitur-fitur ini konsisten dengan literatur klinis yang menyebutkan bahwa hipertensi, usia lanjut, dan penyakit penyerta seperti diabetes merupakan faktor risiko utama penyakit jantung. Hal ini menunjukkan gejala akut memiliki bobot prediktif lebih tinggi dalam dataset ini dibandingkan komorbiditas kronis.

Hasil penelitian ini menunjukkan bahwa :

- a. 1, XGBoost merupakan algoritma yang sangat efektif untuk klasifikasi risiko penyakit jantung dengan data tabular
- b. 2. Hyperparameter tuning secara sistematis mampu meningkatkan performa model secara material, terutama dalam mengurangi kesalahan klasifikasi pada kelas positif (risiko Tinggi)
- c. fitur klinis dan gaya hidup tertentu memiliki dampak signifikan terhadap klasifikasi risiko, yang konsisten dengan pemahaman medis saat ini
- d. penurunan jumlah false negative menjadi sorotan penting karena hal ini berdampak langsung terhadap keputusan klinis yang berbasis pada hasil model.

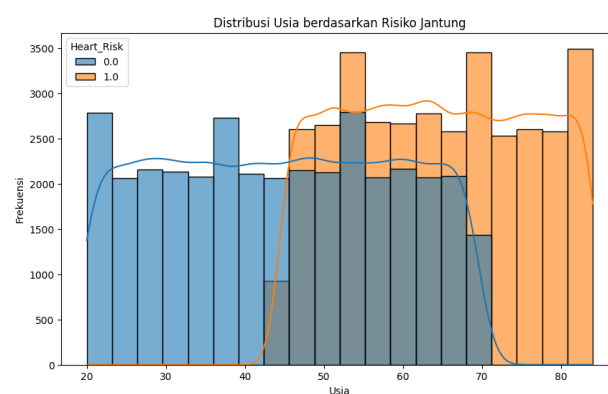


Gambar 1. Correlation Matrix Pearson seluruh Fitur

Gambar diatas merupakan heatmap matriks korelasi Pearson yang menggambarkan hubungan linier antar seluruh fitur dalam dataset prediksi risiko penyakit jantung. Setiap nilai korelasi direpresentasikan melalui skala warna, di mana warna merah menunjukkan korelasi tinggi (mendekati 1), dan warna biru menunjukkan korelasi rendah atau tidak signifikan (mendekati 0). Diagonal utama dari matriks menunjukkan nilai 1.00 karena setiap fitur dibandingkan dengan dirinya sendiri.

Berdasarkan pengamatan terhadap kolom terakhir, yaitu Heart_Risk, ditemukan bahwa sejumlah fitur memiliki korelasi yang relatif tinggi terhadap risiko penyakit jantung. Fitur-fitur tersebut antara lain Chest_Pain, Shortness_of_Breath, Fatigue, Pain_Arms_Jaw_Back, Cold_Sweats_Nausea, dan Swelling, yang masing-masing memiliki korelasi sebesar 0.60. Selain itu, fitur Age tercatat memiliki nilai korelasi tertinggi terhadap target, yaitu sebesar 0.61. Hal ini menunjukkan bahwa fitur-fitur tersebut memiliki kontribusi penting dalam memengaruhi prediksi risiko penyakit jantung, dan dapat dianggap sebagai prediktor utama dalam pembangunan model klasifikasi.

Secara umum, sebagian besar fitur dalam dataset tidak menunjukkan hubungan korelasi yang sangat tinggi satu sama lain (nilai korelasi di bawah 0.4), yang mengindikasikan bahwa tidak terdapat masalah multikolinearitas yang berarti. Oleh karena itu, fitur-fitur yang tersedia dapat digunakan secara bersamaan dalam pelatihan model tanpa perlu melakukan penghapusan fitur akibat redundansi statistik. Visualisasi ini menjadi dasar yang penting dalam tahap eksplorasi data karena membantu dalam mengidentifikasi fitur-fitur relevan dan memastikan bahwa input model tidak saling bertumpang tindih secara linier.



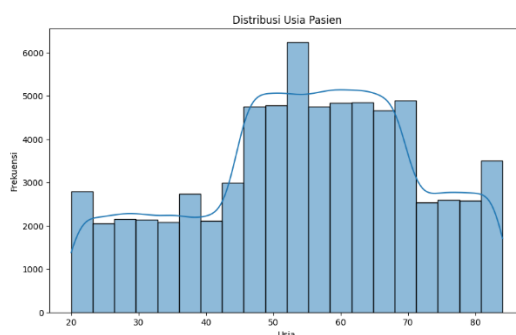
Gambar 2. Histogram Berganda Berdasarkan Risiko Penyakit Jantung

Gambar di atas menunjukkan distribusi usia pasien berdasarkan kategori risiko penyakit jantung, yang diwakili oleh variabel biner Heart_Risk (0 = risiko rendah, 1 = risiko tinggi). Histogram dibagi ke dalam dua kelompok, ditandai dengan dua warna berbeda: biru untuk pasien dengan risiko rendah dan oranye untuk pasien dengan risiko tinggi. Garis lengkung pada masing-masing kelompok menggambarkan estimasi kepadatan distribusi (density plot) yang membantu memahami pola penyebaran data secara halus.

Dari grafik dapat diamati bahwa pasien dengan risiko rendah cenderung berada pada rentang usia yang lebih muda, yaitu antara usia 20 hingga 50 tahun, dengan puncak distribusi yang relatif merata. Sebaliknya, pasien

dengan risiko tinggi umumnya berada pada usia di atas 50 tahun, dengan konsentrasi yang lebih tinggi pada rentang usia 55 hingga 75 tahun. Distribusi pasien risiko tinggi tampak menonjol di usia-usia lanjut, dengan jumlah yang terus meningkat hingga mendekati usia 80 tahun.

Polanya menunjukkan bahwa terdapat korelasi positif antara peningkatan usia dengan risiko penyakit jantung. Dengan kata lain, semakin tua usia seseorang, semakin besar kemungkinan ia termasuk ke dalam kategori risiko tinggi. Temuan visual ini sejalan dengan pengetahuan medis bahwa usia merupakan salah satu faktor risiko utama penyakit jantung. Oleh karena itu, usia menjadi salah satu variabel penting dalam model prediktif yang digunakan untuk mengklasifikasikan risiko penyakit jantung dalam penelitian ini.

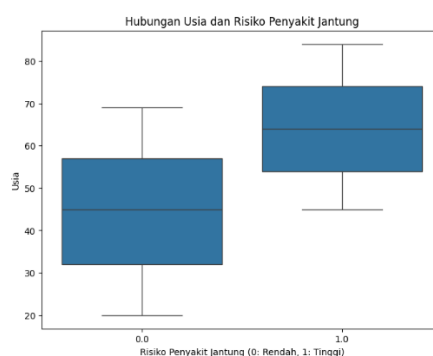


Gambar 3. Density Plot Berdasarkan Resiko Penyakit Jantung

Gambar di atas memperlihatkan distribusi usia seluruh pasien dalam dataset prediksi risiko penyakit jantung. Visualisasi ini disajikan dalam bentuk histogram yang menunjukkan frekuensi jumlah pasien pada berbagai rentang usia, disertai dengan kurva kepadatan (density plot) yang memberikan gambaran pola sebaran data secara lebih halus.

Distribusi usia pasien terlihat cukup beragam dan menyebar luas, mulai dari usia sekitar 20 hingga lebih dari 80 tahun. Puncak distribusi (modus) berada pada rentang usia 50 hingga 60 tahun, dengan jumlah pasien yang sangat dominan, mencapai lebih dari 6.000 orang. Selain itu, terdapat juga konsentrasi yang cukup tinggi pada rentang usia 55–70 tahun, yang mencerminkan populasi dewasa lanjut yang secara umum lebih rentan terhadap penyakit jantung.

Meskipun distribusinya tidak sepenuhnya simetris, pola sebarannya menunjukkan kecenderungan normal dengan sedikit skew ke kanan (right-skewed), mengindikasikan bahwa pasien lansia cenderung lebih banyak dibanding pasien usia muda. Distribusi ini juga menunjukkan bahwa dataset memiliki representasi usia yang memadai untuk analisis prediktif, khususnya dalam konteks penyakit yang prevalensinya meningkat seiring bertambahnya usia.



Gambar 4. Box Plot Usia vs Heart _ Risk

Gambar di atas menyajikan boxplot yang menggambarkan hubungan antara usia pasien dan kategori risiko penyakit jantung (Heart_Risk) dalam dataset. Sumbu horizontal menunjukkan dua kategori risiko (0 = risiko rendah, 1 = risiko tinggi), sedangkan sumbu vertikal menunjukkan usia pasien.

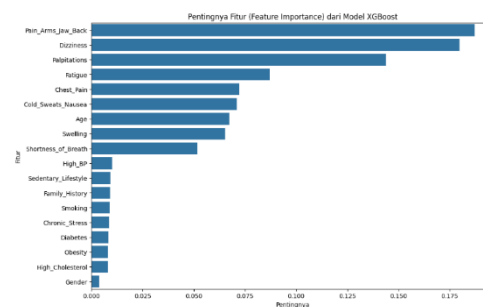
Dari visualisasi ini terlihat bahwa pasien dengan risiko penyakit jantung tinggi (kategori 1) secara signifikan memiliki usia yang lebih tinggi dibandingkan pasien dengan risiko rendah. Median usia pasien risiko tinggi berada di sekitar 65 tahun, dengan rentang interkuartil (IQR) antara kurang lebih 55 hingga 75 tahun. Sebaliknya, pada

kategori risiko rendah, median usia pasien tercatat lebih rendah, yaitu sekitar 45 tahun, dengan IQR antara 35 hingga 60 tahun.

Plot ini juga menunjukkan bahwa penyebaran usia pada kategori risiko tinggi lebih terkonsentrasi pada kelompok lansia, sementara pasien pada kategori risiko rendah mencakup populasi yang lebih muda. Selain itu, tidak ditemukan banyak outlier, yang menandakan distribusi usia relatif stabil dalam masing-masing kelompok.

Secara keseluruhan, boxplot ini mengonfirmasi bahwa usia merupakan salah satu faktor yang sangat membedakan antara kelompok berisiko rendah dan tinggi terhadap penyakit jantung, dan dengan demikian, merupakan salah satu fitur penting dalam proses klasifikasi prediktif.

Performance XGBoost Sebelum menggunakan Hypertuning Grid Search



Gambar 5. Feature Importance sebelum Hypertuning

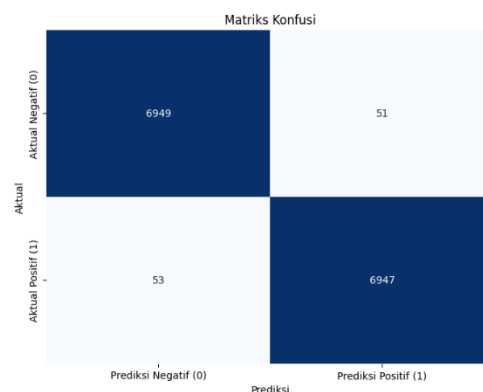
Gambar di atas menunjukkan visualisasi pentingnya fitur (feature importance) yang dihasilkan oleh model XGBoost dalam memprediksi risiko penyakit jantung. Grafik ini menyajikan kontribusi relatif masing-masing fitur terhadap performa model, diukur berdasarkan tingkat pengaruhnya dalam proses pemisahan data selama pembangunan pohon keputusan.

Fitur yang memiliki pengaruh paling besar dalam klasifikasi risiko penyakit jantung adalah Pain_Arms_Jaw_Back, diikuti oleh Dizziness dan Palpitations. Ketiga fitur ini memiliki nilai kepentingan tertinggi, yang menunjukkan bahwa gejala-gejala fisik tersebut sangat relevan dalam membedakan antara pasien dengan dan tanpa risiko jantung.

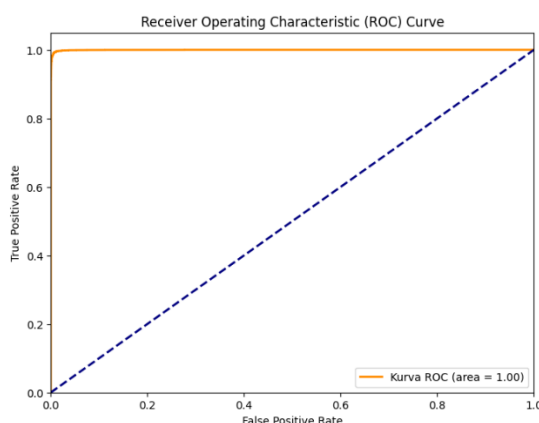
Fitur lain yang juga memiliki kontribusi signifikan adalah Fatigue, Chest_Pain, Cold_Sweats_Nausea, dan Age. Keberadaan Age sebagai salah satu fitur penting memperkuat temuan sebelumnya bahwa usia memainkan peran penting dalam risiko penyakit jantung. Sementara itu, fitur-fitur seperti Swelling, Shortness_of_Breath, dan High_BP juga menunjukkan pengaruh yang moderat terhadap prediksi model.

Di sisi lain, beberapa fitur seperti Gender, High_Cholesterol, Obesity, dan Diabetes memiliki nilai penting yang relatif rendah dalam model ini. Meskipun secara klinis faktor-faktor tersebut dikenal sebagai risiko penyakit jantung, namun kontribusinya dalam konteks dataset dan model XGBoost ini tidak sebesar fitur gejala langsung seperti nyeri atau kelelahan.

Secara keseluruhan, visualisasi ini memberikan wawasan mengenai variabel-variabel yang paling berpengaruh dalam membentuk keputusan model, yang sangat berguna untuk interpretasi medis dan validasi klinis. Hasil ini juga dapat digunakan untuk menyederhanakan model dengan tetap mempertahankan akurasi melalui seleksi fitur yang relevan.



Gambar 6. Confusion Matrix Sebelum Hyperparameter Tuning



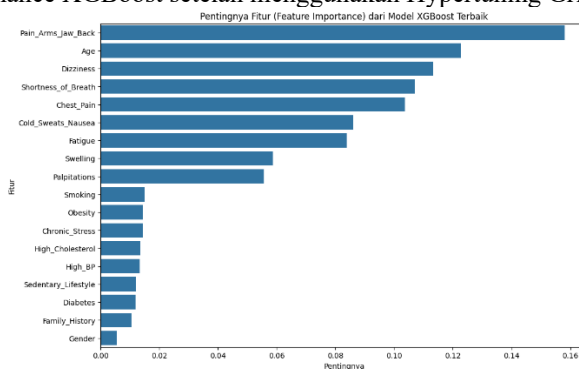
Gambar 7. ROC Curve Sebelum Hyperparameter Tuning

Gambar di atas merupakan matriks konfusi dari hasil evaluasi model XGBoost yang telah dioptimalkan menggunakan GridSearchCV pada data pengujian. Matriks ini menunjukkan jumlah prediksi yang benar dan salah berdasarkan dua kelas: risiko penyakit jantung rendah (0) dan tinggi (1).

Berdasarkan hasil tersebut, model berhasil mengklasifikasikan 6.949 pasien dengan risiko rendah secara benar (True Negative), dan 6.947 pasien dengan risiko tinggi juga terklasifikasi dengan benar (True Positive). Sementara itu, terdapat 51 kasus False Positive, yaitu pasien yang sebenarnya tidak berisiko namun diprediksi berisiko tinggi oleh model. Sebaliknya, 53 kasus False Negative terjadi ketika pasien yang sebenarnya berisiko tinggi diprediksi sebagai tidak berisiko.

Secara keseluruhan, hasil ini menunjukkan bahwa model memiliki kinerja klasifikasi yang sangat baik, dengan jumlah kesalahan prediksi yang sangat kecil dibandingkan total prediksi (hanya sekitar 0,74% kesalahan dari total 14.000 data uji). Proporsi True Positive Rate (Recall) yang tinggi mengindikasikan bahwa model efektif dalam mengidentifikasi pasien yang benar-benar berisiko tinggi. Selain itu, rendahnya jumlah False Negative menjadi sangat penting dalam konteks medis, karena kegagalan mendeteksi pasien berisiko tinggi dapat berdampak serius terhadap intervensi klinis.

Performance XGBoost setelah menggunakan Hypertuning Grid Search



Gambar 8. Feature Importance setelah Menggunakan Hypertuning Grid Search

Gambar di atas menampilkan visualisasi tingkat kepentingan fitur (feature importance) dari model XGBoost terbaik yang telah dioptimalkan melalui hyperparameter tuning menggunakan GridSearchCV. Grafik ini menunjukkan kontribusi relatif masing-masing fitur dalam mempengaruhi keputusan klasifikasi model terhadap risiko penyakit jantung.

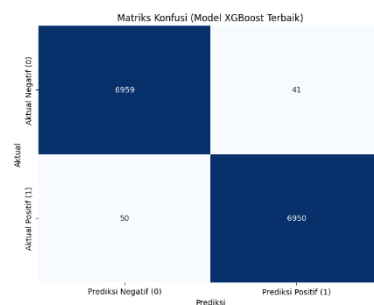
Fitur yang paling dominan dalam model ini adalah Pain_Arms_Jaw_Back, yang memiliki nilai kepentingan tertinggi dibandingkan fitur lainnya. Ini menegaskan bahwa nyeri pada lengan, rahang, atau punggung merupakan indikator yang sangat signifikan dalam mendeteksi risiko penyakit jantung. Di posisi kedua, Age (usia) juga muncul sebagai faktor prediktif yang sangat penting, sejalan dengan temuan epidemiologis bahwa risiko penyakit jantung meningkat seiring bertambahnya usia.

Fitur lain yang juga memiliki pengaruh besar meliputi Dizziness (pusing), Shortness_of_Breath (sesak napas), dan Chest_Pain (nyeri dada), yang semuanya merupakan gejala klinis utama penyakit jantung. Selanjutnya,

fitur seperti Cold Sweats Nausea, Fatigue (kelelahan), dan Swelling (pembengkakan) juga menunjukkan kontribusi moderat terhadap kinerja model.

Sementara itu, fitur-fitur seperti Gender, Family History, dan Diabetes menunjukkan tingkat kepentingan yang relatif rendah. Hal ini mengindikasikan bahwa meskipun faktor-faktor tersebut memiliki relevansi secara medis, dalam konteks dataset ini kontribusinya terhadap prediksi risiko penyakit jantung tidak sebesar gejala klinis langsung atau variabel usia.

Secara keseluruhan, grafik ini memberikan wawasan penting mengenai fitur-fitur yang paling berpengaruh dalam model prediksi dan dapat dijadikan acuan dalam pengembangan sistem pendukung keputusan medis berbasis data, terutama untuk fokus pada gejala dan faktor demografis yang terbukti paling menentukan dalam prediksi risiko penyakit jantung.



Gambar 9. Confusion Matrix Setelah Hyperparameter Tuning

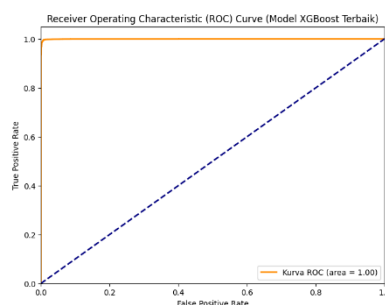
Gambar tersebut menunjukkan matriks konfusi dari model XGBoost terbaik yang telah dioptimalkan menggunakan teknik Grid Search. Matriks ini merepresentasikan performa prediksi model pada data uji, dengan empat elemen utama: True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP).

Dari hasil yang ditampilkan, model berhasil mengklasifikasikan:

- 6.959 data negatif secara benar (TN), yaitu pasien yang benar-benar tidak berisiko tinggi dan diprediksi sebagai tidak berisiko tinggi.
- 6.950 data positif secara benar (TP), yaitu pasien yang benar-benar berisiko tinggi dan diprediksi dengan tepat.
- Terdapat 41 kasus False Positive (FP), di mana model memprediksi pasien sebagai berisiko tinggi padahal kenyataannya tidak.
- Terdapat 50 kasus False Negative (FN), di mana model gagal mengidentifikasi pasien yang sebenarnya berisiko tinggi.

Kinerja model ini menunjukkan tingkat akurasi yang sangat tinggi, ditandai dengan jumlah TP dan TN yang jauh lebih besar dibanding FP dan FN. Namun, jumlah FN (50) tetap menjadi fokus penting karena pada konteks klinis, kegagalan dalam mendeteksi pasien berisiko tinggi dapat berakibat serius. Meskipun angkanya kecil secara absolut, penurunan jumlah FN dari baseline model sebelumnya menandakan adanya peningkatan keandalan model hasil tuning dalam deteksi risiko jantung.

Matriks ini juga mendukung nilai ROC AUC dan akurasi tinggi yang diperoleh sebelumnya, memperkuat bukti bahwa model XGBoost yang telah dioptimalkan tidak hanya akurat secara statistik, tetapi juga relevan secara klinis dalam mendeteksi pasien dengan risiko penyakit jantung.



Gambar 10. ROC Curve Setelah Hyperparameter Tuning

Gambar di atas menunjukkan kurva ROC (Receiver Operating Characteristic) untuk model XGBoost terbaik yang telah dioptimalkan melalui proses hyperparameter tuning. Kurva ini memvisualisasikan kemampuan diskriminatif model dalam membedakan antara dua kelas target, yaitu pasien dengan dan tanpa risiko tinggi penyakit jantung.

Sumbu horizontal merepresentasikan False Positive Rate (FPR), sedangkan sumbu vertikal merepresentasikan True Positive Rate (TPR). Kurva ROC model ditunjukkan dengan garis berwarna oranye, sementara garis biru putus-putus diagonal menunjukkan baseline model acak (random classifier), yaitu model yang menebak secara acak tanpa kemampuan klasifikasi yang berarti.

Kurva oranye berada sangat dekat dengan sudut kiri atas grafik, yang merupakan indikasi kuat bahwa model mampu mencapai sensitivitas yang tinggi dengan tingkat kesalahan yang sangat rendah. Nilai AUC (Area Under Curve) tercatat sebesar 1.00, yang merupakan nilai sempurna. Hal ini menunjukkan bahwa model memiliki kemampuan hampir sempurna dalam membedakan antara kelas positif (risiko tinggi) dan negatif (risiko rendah) di berbagai threshold prediksi.

Secara keseluruhan, grafik ini memberikan bukti kuat bahwa model XGBoost yang telah dioptimalkan tidak hanya memiliki performa prediksi yang tinggi secara kuantitatif, tetapi juga sangat andal secara klinis, karena kemampuannya dalam meminimalkan kesalahan klasifikasi yang berisiko.

4. KESIMPULAN

Penelitian ini memberikan kontribusi orisinal dalam bidang medical data mining dengan menunjukkan bahwa optimasi sistematis hiperparameter XGBoost melalui Grid Search dapat menghasilkan peningkatan performas yang bermakna secara klinis, meskipun pada dataset yang sudah berkinerja sangat tinggi. Temuan utama dalam penelitian ini yaitu pengurangan *false negatives* sebesar 1,2 kasus per 14.000 prediksi, yang menegaskan bahwa perbaikan numerik yang kecil dapat memiliki dampak nyata dalam konteks deteksi dini penyakit jantung.

Penelitian ini juga memperkuat validitas XGBoost sebagai alat prediktif andal untuk data medis tabular, sekaligus menawarkan kerangka kerja yang dapat direplikasi untuk optimasi model berbasis gradient boosting dalam aplikasi kesehatan.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia atas dukungan pendanaan melalui Program Hibah Penelitian Dasar Pemula (PDP) Tahun Anggaran 2025. Penulis juga menyampaikan apresiasi kepada institusi dan semua pihak yang telah memberikan kontribusi dan dukungan selama proses penelitian ini berlangsung.

DAFTAR PUSTAKA

- [1] R. R. Aulia, F. H. Lubis, and L. S. Harahap, "Penerapan Multi-Layer Perceptron untuk Prediksi Durasi Tidur Berdasarkan Faktor Kebiasaan Harian," *Journal of Multidisciplinary Inquiry in Science Technology and Educational Research*, vol. 2, no. 1, pp. 20–30, 2025, doi: 10.32672/mister.v2i1.2326.
- [2] U. A. Syam, I. Irdayanti, I. Magfirrah, B. Sartono, and A. R. Firdawanti, "Evaluasi Kinerja Model Random Forest dan LightGBM untuk Klasifikasi Status Imunisasi Hepatitis B (HB-0) pada Balita," *Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 1, pp. 1–8, Apr. 2025, doi: 10.37905/euler.v13i1.29762.
- [3] F. V. Ongkosianbhadra and C. C. Lestari, "Pengembangan Model Prediksi Risiko Hipertensi Menggunakan Algoritma Gradient Boosting Decision Tree Yang Dioptimalkan," *Jurnal Informatika dan Sistem Informasi*, vol. 9, no. 2, pp. 90–99, Dec. 2023, doi: 10.37715/juisi.v9i2.4403.
- [4] G. Abdurrahman, H. Oktavianto, and M. Sintawati, "Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridsearch dan Random Search Pada Klasifikasi Penyakit Diabetes," *Informatics Journal*, vol. 7, no. 3, 2022, doi : <https://doi.org/10.19184/isj.v7i3.35441>.
- [5] V. Marcella Angela Simalango and W. Franciska Senjaya, "Evaluasi Algoritma Decision Tree dan Random Forest serta Efektivitas Feature Selection dalam Memprediksi Kesehatan Mental," *Jurnal Strategi*, vol 7 no. 1, 2025.
- [6] "Implementasi Data Mining Dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode Random Forest Dan Xgboost," *Jurnal Ilmiah Komputasi*, vol. 23, no. 1, Mar. 2024, doi: 10.32409/jikstik.23.1.3507.
- [7] S. Handayani, D. Toresa, T. Informatika, F. Ilmu Komputer, and U. Lancang Kuning, "Peningkatan

-
- Performa Model Gradient Boosting dalam Klasifikasi Stroke Melalui Optimasi Grid Search”, *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 14, no.3, 2024, doi: <https://doi.org/10.37859/jf.v14i3.7893>
- [8] M. RoisS, “PERBANDINGAN KINERJA SUPPORT VECTOR MACHINE DALAM KLASIFIKASI OBESITAS DENGAN PENDEKATAN KERNEL LINEAR DAN RADIAL BASIS FUNCTION,” *JURNAL DEVICE*, vol. 15, no. 1, pp. 14–23, 2025.
- [9] S. Mujiyono, U. P. Sanjaya, I. S. Wibisono, and H. Setyowati, “Prediksi Fluktuasi Berat Badan Berdasarkan Pola Hidup Menggunakan Model XGBoost dan Deep Learning,” *Jurnal Algoritma*, vol. 22, no. 1, pp. 221–233, May 2025, doi: 10.33364/algoritma/v.22-1.2253.
- [10] H. Karmila and S. Yuliyatini, “SYSTEMATIC LITERATURE REVIEW: EFFECTIVE DATA NORMALIZATION METHOD IN DETECTING DIABETES USING MACHINE LEARNING (K-NEAREST NEIGHBORS ALGORITHM) SYSTEMATIC LITERATURE REVIEW: METODE NORMALISASI DATA YANG EFEKTIF PADA DETEKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE LEARNING (ALGORITMA K-NEAREST NEIGHBORS),” *Journal of Scientech Research and Development*, vol. 6, no. 1, 2024, [Online]. Available: <https://idm.or.id/JSCR/inde>
- [11] A. Muliawati and H. Nurramdhani Irmanda, “Penerapan Borderline-SMOTE dan Grid Search pada Bagging-SVM untuk Klasifikasi Penyakit Diabetes,” *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasi*, 2022.
- [12] M. A. Saputra and T. Sugihartono, “Evaluasi Kinerja Model LSTM Untuk Prediksi Risiko Penyakit Jantung Menggunakan Dataset,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 7, pp. 1823–1833, Jul. 2025, doi: 10.52436/1.jpti.821.
- [13] A. Faqih and T. Sugihartono, “Perbandingan Algoritma XGBoost dan LSTM dalam Prediksi Harga Saham Tesla Menggunakan Data Tahun 2025,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 6, pp. 1563–1573, Jun. 2025, doi: 10.52436/1.jpti.836.